Współczynniki korelacji i determinacji wielorakiej

Rozważmy wektor losowy $(Y, X_1,...,X_k)^T$ o wektorze wartości oczekiwanych $(m_y, m_1,...,m_k)^T$ macierzy kowariancji w postaci blokowej $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_y^2 & \boldsymbol{\sigma}_{yX} \\ \boldsymbol{\sigma}_{Xy} & \boldsymbol{\Sigma}_{XX} \end{bmatrix}$, gdzie $\boldsymbol{\sigma}_{yX} = \boldsymbol{\sigma}_{Xy}^{\mathrm{T}}$.

Problem. Szukamy miary zależności zmiennej losowej Y od wektora $\mathbf{X} = (X_1,...,X_k)^{\mathrm{T}}$.

Poszukujemy liniowej kombinacji $\sum_{i=1}^{k} \beta_i X_i = \langle \boldsymbol{\beta}, \mathbf{X} \rangle = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$, która ma największy współczynnik korelacji ze zmienną Y.

$$\rho(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) = \frac{\operatorname{cov}(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\sqrt{V(Y)V(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}} = \frac{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\sigma}_{Xy}}{\sigma_y\sqrt{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}\boldsymbol{\beta}}}.$$

Ponieważ współczynnik korelacji jest niezmienniczy ze względu na zmiany położenia i skali wektora $\mathbf{X}$ i zmiennej Y, możemy bez straty ogólności, że

- wszystkie rozważane zmienne są scentrowane (tzn. mają zerowe wartości oczekiwane)
- $V(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}\boldsymbol{\beta} = 1$.
- $\sigma_y^2 = 1$ i .

Przy powyższych ograniczeniach współczynnik korelacji $\rho(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$ przyjmuje wartość największą gdy licznik $\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\sigma}_{Xy}$ przyjmuje wartość największą przy powyższych ograniczeniach.

Tworzymy funkcje Lagrange'a $L(\boldsymbol{\beta}, \lambda) = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\sigma}_{Xy} - \frac{1}{2}\lambda(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}\boldsymbol{\beta} - 1)$.

WK $$\begin{cases} \dfrac{\partial}{\partial \boldsymbol{\beta}} L(\boldsymbol{\beta}, \lambda) = \boldsymbol{\sigma}_{Xy} - \lambda\boldsymbol{\Sigma}_{XX}\boldsymbol{\beta} = \mathbf{0} \\ \dfrac{\partial L(\boldsymbol{\beta}, \lambda)}{\partial \lambda} = -\frac{1}{2}(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}\boldsymbol{\beta} - 1) = 0 \end{cases} \Leftrightarrow \begin{cases} \boldsymbol{\beta} = \frac{1}{\lambda}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\sigma}_{Xy} \\ \lambda^2 = \boldsymbol{\sigma}_{Xy}^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\sigma}_{Xy} \end{cases}.$$

Stąd $\lambda = \sqrt{\boldsymbol{\sigma}_{Xy}^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\sigma}_{Xy}}$ i $\boldsymbol{\beta} = \dfrac{\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\sigma}_{Xy}}{\sqrt{\boldsymbol{\sigma}_{Xy}^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\sigma}_{Xy}}}$.

Maksymalny współczynnik korelacji , czyli populacyjny współczynnik korelacji wielorakiej jest więc równy

$$\rho_{yX} = \frac{\sqrt{\boldsymbol{\sigma}_{Xy}^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\sigma}_{Xy}}}{\sigma_y} \text{ a jego kwadrat } \rho_{yX}^2 = \frac{\boldsymbol{\sigma}_{Xy}^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\sigma}_{Xy}}{\sigma_y^2}.$$

Jeśli w powyższych wzorach zastąpimy momenty teoretyczne ich estymatorami tzn. momentami próbkowymi, to otrzymamy wzory na próbkowy współczynnik korelacji wielorakiej i jego kwadrat próbkowy współczynnik determinacji.

Estymatorem macierzy kowariancyjnej $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_y^2 & \boldsymbol{\sigma}_{yX} \\ \boldsymbol{\sigma}_{Xy} & \boldsymbol{\Sigma}_{XX} \end{bmatrix}$ jest próbkowa macierz kowariancji

$$\mathbf{S} = \begin{bmatrix} s_y^2 & \mathbf{s}_{yX} \\ \mathbf{s}_{Xy} & \mathbf{S}_{XX} \end{bmatrix}.$$

Tak więc $R_{yX} = \frac{\sqrt{\mathbf{s}_{Xy}^{\mathrm{T}} \mathbf{S}_{XX}^{-1} \mathbf{s}_{Xy}}}{s_y}$ i $R_{yX}^2 = \frac{\mathbf{s}_{Xy}^{\mathrm{T}} \mathbf{S}_{XX}^{-1} \mathbf{s}_{Xy}}{s_y^2}$ są próbkowymi współczynnikami korelacji i determinacji wielokrotnej

Korzystając z wzoru $\det \begin{bmatrix} a_{11} & \mathbf{a}_{21}^{\mathrm{T}} \\ \mathbf{a}_{21} & \mathbf{A}_{22} \end{bmatrix} = \det \mathbf{A}_{22} (a_{11} - \mathbf{a}_{21}^{\mathrm{T}} \mathbf{A}_{22}^{-1} \mathbf{a}_{21})$ możemy zapisać

$$\det(\boldsymbol{\Sigma}) = \det \begin{bmatrix} \sigma_y^2 & \boldsymbol{\sigma}_{yX} \\ \boldsymbol{\sigma}_{Xy} & \boldsymbol{\Sigma}_{XX} \end{bmatrix} = \det(\boldsymbol{\Sigma}_{XX}) \left(\sigma_y^2 - \boldsymbol{\sigma}_{yX} \boldsymbol{\Sigma}_{XX}^{-1} \boldsymbol{\sigma}_{Xy} \right) \text{ i stąd}$$

Populacyjny współczynnik determinacji jest określony wzorem

$$\rho_{yX}^2 = 1 - \frac{\det(\boldsymbol{\Sigma})}{\sigma_y^2 \det(\boldsymbol{\Sigma}_{XX})},$$

a jego próbkowym estymator jest równy

$$R_{yX}^2 = 1 - \frac{\det(\mathbf{S})}{s_y^2 \det(\mathbf{S}_{XX})}.$$

Rozkład tego współczynnika z *n*-elementowej próby z wielowymiarowego rozkładu normalnego podał Fisher (1928). Wishart(1931) wyznaczył momenty tego rozkładu. W szczególności dla $\rho_{yX} = 0$ (tzn. gdy *Y* nie zależy od $(X_1,...,X_k)$

$$E(\hat{R}_{yX}^2) = \frac{k}{n-1},$$

co oznacza że estymator $\hat{R}_{yX}^2$ jest obciążony. Aby skorygować to obciążenie odejmiemy od estymatora $\frac{k}{n-1}$. Ta korekta spowoduje jednak, że estymator $\hat{R}_{yX}^2 - \frac{k}{n-1}$ będzie przyjmował zbyt małe wartości dla dużych wartości $\hat{R}_{yX}^2$. Skalujemy więc tak $\hat{R}_{yX}^2 - \frac{k}{n-1}$ aby dla $\hat{R}_{yX}^2 = 1$ również przeskalowany

skorygowany współczynnik przyjął wartość 1. Definiujemy więc ostatecznie skorygowany (adjusted R^2) jako

$$\hat{R}^2_{yX,adj} = \frac{\hat{R}^2_{yX} - \frac{k}{n-1}}{1-\frac{k}{n-1}} = \frac{(n-1)\hat{R}^2_{yX} - k}{n-1-k}.$$

Współczynniki korelacji cząstkowej

Rozważmy wektor losowy $(X_1,...,X_k)^T$ o macierzy kowariancyjnej postaci. Miarą liniowej zależności pomiędzy zmiennymi X_i oraz X_j jest współczynnik korelacji Pearsona

$$\rho_{ij} = \frac{\operatorname{cov}(X_i, X_j)}{\sqrt{V(X_i)V(X_j)}}.$$

Współczynnik korelacji cząstkowej będzie marą liniowej zależności pomiędzy zmiennymi X_i oraz X_j po wyeliminowaniu wpływu pozostałych zmiennych na obie zmienne X_i oraz X_j.

Rozważmy nowe zmienne

$$X_i^* = X_i - b_{i0} - \sum_{\substack{q=1 \\ q\neq i,j}}^{k} b_{iq}X_q, \qquad X_j^* = X_j - b_{j0} - \sum_{\substack{q=1 \\ q\neq i,j}}^{k} b_{jq}X_q,$$

przy czym $E(X_i^*)^2 = \min$ i $E(X_j^*)^2 = \min$. Można powiedzieć że są to zmienne resztowe powstałe przez odjecie od oryginalnej zmiennej jej najlepszej średniokwadratowej aproksymacji afiniczną funkcją pozostałych zmiennych losowych (różnych od X_i oraz X_j).

Współczynnik korelacji

$$\kappa_{ij} = \frac{\operatorname{cov}(X_i^*, X_j^*)}{\sqrt{V(X_i^*)V(X_{j\,j}^*)}}$$

pomiędzy tymi nowymi zmiennymi jest nazywany współczynnikiem korelacji cząstkowej pomiędzy zmiennymi X_i oraz X_j.

Można pokazać, że $\kappa_{ij} = \frac{-P_{ij}}{\sqrt{P_{ii}P_{jj}}}$, gdzie P_{ij} jest dopełnieniem algebraicznym elementu ρ_{ij} macierzy korelacyjnej

$$P = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{21} & 1 & \cdots & \rho_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1} & \rho_{k2} & \cdots & 1 \end{bmatrix}.$$